

SPE 29756

A Cost-Effective Statistical Screening Method to Detect Oilfield Brine Contamination

Nancy Alyanak, C.P.G., Consultant, Spring, Texas, James T. Grace, C.P.G., Consultant, Spring, Texas, Michael D. Campbell, P.G., P.H., United Resources Int., Houston, Texas

Copyright 1995, Society of Petroleum Engineers, Inc.

This paper was prepared for presentation at the SPE/EPA Exploration & Production Environmental Conference held in Houston, TX, U.S.A., 27-29 March 1995.

This paper was selected for presentation by an SPE Program Committee following review of Information contained in an abstract submitted by the author(s). Contents of the paper, as presented, have not been reviewed by the Society of Petrolem Engineers and are subject to correction by the author(s). The material, as presented, does not necessarily reflect any position of the Society of Petroleum Engineers, its officers, or members. Papers presented at SPE meetings are subject to publication review by Editorial Committees of Petroleum Engineers. Permission to copy is restricted to an abstract of not more than 300 words. *Rustrations may* not be copied. The abstract should contain conspicuous acknowledgment of where and by whom the paper is presented. Write Librarian, SPE, P.O. Box 833836, Richardson, TX 75083-3838, U.S.A., Telex, 163245 SPEUT.

ABSTRACT

A statistical screening method has been developed using Tolerance Limits for barium (Ba⁺²) to identify contamination of a fresh-water aquifer by oilfield brines. The method requires an understanding of the local hydrochemistry of oilfield brines, inexpensive, publicly available hydrochemical data, a single sample analysis from the suspect well and the application of a simple statistical procedure. While this method may not provide absolute evidence of oilfield brine contamination of a fresh-water aquifer, it does identify conditions where brine contamination is a strong probability over other possible sources of chlorides.

INTRODUCTION

As a conservative contaminant, chlorides move through the hydrologic cycle as a result of physical processes. Objectionable because of undesirable physical effects (taste, odor, corrosion), chlorides survive most processes which remove other ions from naturally occurring waters. While normally not a

References and illustrations at end of paper

health hazard, high chloride levels do affect the aesthetic quality of water and can severely limit the use of a water supply. As a result, finding the cause of increasing chloride concentrations in a fresh-water aquifer has long been an environmental concern.

Chloride contamination can result from a number of sources: industrial wastes, road de-icing, septic tanks, oilfield brines, seawater and evaporite rocks to name a few. With a 400 mg/L taste threshold for most people¹, chlorides are easily detected. However, as regulated by the Safe Drinking Water Act, chloride (Cl⁻) concentrations are of concern when they exceed the SDWA's secondary contaminant limits of 250 mg/L.

Determining the source of increasing chloride levels can be difficult. Using the concentration of barium (Ba⁺²) as an indicator variable, a statistical screening method has been developed which can be used to identify the contamination of a fresh-water aquifer by oilfield brines. In the current cost-sensitive economic climate it is particularly important to find cost-effective methods to identify oilfield brine contamination. The screening method requires an understanding of the local hydrochemistry of oilfield brines, a source for background hydrochemical data,

a single sample analysis from a suspect well and the application of a simple statistical procedure -- Tolerance Limits.

The procedures used for the calculation of Tolerance Limits and their application as a screening method are demonstrated using an example dataset from Wood County, Texas. While the calculation of a Tolerance Limit is not difficult, it can be time consuming for large datasets. A new freeware computer program, "A Ground Water Information Tracking System with Statistical Analysis Capability" (GRITS/STAT) developed by the United States Environmental Protection Agency² (USEPA) can quickly and accurately determine Tolerance Limits for an indicator variable. Using GRITS/STAT, comparison of the concentrations of the indicator variable from numerous well-sites to the established Tolerance Limit can be rapidly accomplished.

While the screening method described in this paper may not provide absolute evidence of oilfield brine contamination of a fresh-water aquifer, it does provide initial evidence for identifying conditions where brine contamination is a strong probability over other possible sources of chlorides.

METHODOLOGY

The Tolerance Limit

. . . .

Tolerance Limits are calculated using sample data from the background population. Basic information about the background population must be determined before choosing any statistical method. Tolerance Limits are not always the appropriate method to choose.

A background population is defined as containing all possible values of a specified variable. For the example dataset in this paper, the background population is the naturally occurring Ba⁺² concentrations measured in the fresh water aquifer(s) of Wood County, Texas.

The pattern of all the possible values a variable can assume constitutes the population distribution. Knowing the pattern or shape of the distribution is important when choosing a statistical method because the distribution type influences the choice and application of a statistical procedure. Most common statistical methods require the population fit a normal distribution, the familiar bell-curve shape. For normally distributed populations, many statistical methods are available to determine the likelihood a sample was taken from a particular population or some other population.

Determining if the measurement of a characteristic belongs to one population or another is the basis for the application of statistical methods within environmental investigations. Tolerance Limits are used to detect a difference between a background population and a sample value possibly contaminated by an objectionable substance. As applied to the screening method. Tolerance Limits test the hypothesis the ionic concentration of a substance in water samples taken from a normal fresh water supply and the ionic concentration of the substance in a suspect water sample are part of the same population. A value or limit is calculated such that all measured values which exceed the limit belong to a population other than the background population. If the value of a sample falls within the predetermined limits, the sample is assumed to have been taken from the normal background population. Therefore, no evidence of contamination would exist.

Tolerance Limits establish a range of concentrations which contain a specified proportion of the chosen population with a stipulated confidence (probability) the range actually contains the desired concentration range. The proportion of the population included within the concentration range is called the coverage. The probability the Tolerance Limit includes the desired coverage is referred to as the confidence factor, also called the tolerance coefficient.

Parametric Tolerance Limits, used herein, are based upon the assumption the background (or parent)

population can be described by a normal distribution. Nonparameteric Tolerance Limits exist but usually prove impractical because of the large number of observations required to provide acceptable levels of coverage, tolerance coefficient and power.

Calculating The Tolerance Limit

Tolerance Limits are simple to calculate. The average background value, the standard deviation of the background values, and a tolerance factor from a statistical table are all that is necessary to calculate Tolerance Limits for a specified population. A tolerance interval is bounded by an upper limit and a lower limit. However, with the exception of pH values, contamination is usually indicated by presence of excessive concentrations of a contaminant. Therefore, the larger limit, called the upper Tolerance Limit, is of importance in most investigations.

For most situations, the upper Tolerance Limit is defined using ground-water samples from background wells or published data. With three or more background samples the Tolerance Limit can be calculated. First the mean, \overline{X} , and the standard deviation, S, are calculated from the values of the background samples. Next, the upper Tolerance Limit is calculated using the equation:

$$TL_{Upper} = \bar{X} + KS$$
(1)

where K is a dimensionless number dependent upon the proportion of the population to be included within the calculated limits and the number of samples in the background. For various sample sizes (n), values for K can be obtained from statistical tables (see Table 1). The tolerance factors in Table 1 provide at least 95% coverage. Tolerance factors are also available for an average of 95% coverage. Factors with at least 95% coverage are recommended for detection monitoring³.

After the upper Tolerance Limit is calculated, watersample values are compared to the upper Tolerance Limit value. Statistically significant evidence of contamination occurs for each water sample value which exceeds the upper Tolerance Limit.

Tolerance Limits have an advantage over other statistical tests - only one analysis from the suspected well needs to be acquired. Other statistical methods commonly used in ground-water monitoring require multiple samples from a suspect well (ANOVA, CABF t-test, etc.) or require a detailed knowledge of future sampling plans (prediction limits). By requiring only one sample the cost is greatly reduced.

Tolerance Limit Use In Ground-Water Monitoring Programs

Tolerance Limits in ground-water monitoring are discussed by Gibbons⁴ and in USEPA guidance documents^{5,3}. The USEPA guidance documents are applicable because the detection monitoring phase of RCRA addresses the same type of problems found in identifying various contaminants in non-RCRA situations. In these Guidance Documents the USEPA presents recommended procedures for using Tolerance Limits for the detection of contamination:

- Background population should not show a high degree of spatial variation.
- Outliers should be identified in the background samples.
- The background samples should be independent and normally distributed or transformed into a normal distribution.
- At least 95% coverage and 95% tolerance coefficient is recommended for a detection screening program.
- A sample size of eight or more background observations is recommended. As little as four can be used.

The Tolerance Limit method requires choosing a value for coverage and a value for confidence. The coverage sets the false positive rate. For example, if we chose 99% coverage (P) with 95% confidence (Y), the calculated Tolerance Limit would produce a 95% probability of getting a 1% false positive rate.

Barium (Ba⁺²) As An Indicator Variable

Ideally, a substance should be present only if contamination has occurred. Chlorides occur naturally in ground water and are derived from many sources. Further, the hydrochemical behavior of chlorides within the ground water is not influenced by the source of chloride. To determine if rising chloride concentrations in a fresh-water supply is a result of one source or another, other constituents must be measured.

Based upon the contrast between the ionic concentration of barium (Ba⁺²) in near subsurface fresh waters (low) to that found in oilfield brines (high), barium can act as a indicator variable for the presence of oilfield brine contamination of a freshwater aquifer. In the shallow subsurface barium reacts with sulfate to form barite (BaSO₄). Very nearly insoluble, barite is a common mineral with a solubility product close to 10⁻¹⁰. According to J. D. Hem¹ (pg.135-137), barite solubility equilibria likely control the concentration of barium in naturally occurring waters. Based upon this assumption, a natural water Ba⁺² concentration range of 0.14 to 0.014 mg/L was predicted - a fairly narrow range.

If the concentration of barium in natural water is controlled by the availability of the element in the environment one would expect a wide range between the upper and lower concentrations within natural water to occur. Two citations support Hem's conclusion on controls of the barium concentration in natural water. Durfor and Beckers indicate the median concentration of barium in public water supplies is 0.043 mg/L. Durum and Haffty determined the median concentration of barium in larger rivers in North America as 0.045 mg/L. Both values are in

agreement with Hem's predictions and are considerably below the U.S. mandatory limit of 1.0 mg/L for any public drinking water supply.

In comparison to natural fresh water aquifers, much higher barium ion concentrations are typically found in oilfield brines. Collins lists average Ba⁺² concentrations by geologic age for producing basins in the United States. As an example, he found Smackover brines average 23 mg/L and Tertiary brines average 73 mg/L. Rittenhouse, et al. provides additional brine analysis information. Other analyses of produced water from fields in Texas show barium concentrations approaching 200 mg/L.

Under the control of barium sulfate solubility equilibria, the number of sources of barium in naturally occurring water is limited. Common sources for increased chloride levels typically do not increase barium concentrations. For example, the average concentration of barium in sea water is 0.03 mg/L^s. Therefore, increased Ba⁺² concentrations, in addition to increased chlorides in a fresh water aquifer can indicate oilfield brine contamination of the aquifer.

Publicly Available Background Data

When calculating Tolerance Limits, gathering and analyzing background samples can add considerable expense. While this maybe necessary at times, publicly available background data can be easily obtained at low cost. Inexpensive hydrochemical and hydrogeological data from the U.S. Geological Survey's WATSTOR Data File can be used to determine background concentrations. Within Texas, data can be acquired through the Texas Water Development Board (TWDB). The Texas Water Development Board's Ground Water Data System contains data from the WATSTOR Data File and data from a variety of other sources. The default data format is ASCII text, character delineated and supplied on a MS-DOS disk.

For Texas, the database contains chemical analyses and water levels from fresh water private, municipal,

agricultural and industrial wells across the state. Each well has an identifying number, latitude, longitude, owner, use and construction details in addition to the chemical analysis and water level data. Some wells contain analyses for over one hundred different items. Other wells have only a few items.

Wood County, Texas - An Example Dataset

Wood County, Texas is located within the East Texas Salt Basin and has been heavily involved in the exploration for oil and gas. Across the county the TWDB's Ground Water Data System maintains records for 413 water wells. Using data from 1988 to present, 20 wells contained analyses for the ionic concentration of barium (Figure 1). For each of the 20 wells, basic information about the well and the Ba⁺² concentration in ug/L and the Cl concentration in mg/L are shown in Table 2.

The first step in a Tolerance Limit calculation is to determine the distribution of the data. Natural data, such as the Ba⁺² concentration in fresh-water aquifers, typically follow a log-normal distribution^{10,3}. As a result, a current USEPA Statistical Guidance Document³ suggests starting with the assumption the data is log-normal. In a log-normal distribution the natural logarithm of the original data is normally distributed, rather than the original data itself.

A probability plot, using the natural logarithm for each value in the Wood County set of data, confirmed the data fit a log-normal distribution (Figure 2). A probability plot is a standard graphical method discussed in many texts as well as the Guidance Documents^{3,5}. If a set of data is normally distributed, the data values will fall along a straight-line on a probability plot. When the natural logarithm for each value of a set of data falls along a straight line on a probability plot, the data is considered to be lognormally distributed.

As a result of the probability plot (Figure 2), the natural logarithm of each Ba⁺² concentration is used to calculate the natural logarithm of the upper

Tolerance Limit. Therefore, for use in comparisons with suspect ground-water samples, the natural logarithm of the upper Tolerance Limit has to be converted from its logarithmic form to its original form using:

$$TL_{upper} = e^{\ln(TL_{upper})}$$
 (2)

The mean of the natural log-transformed data is \overline{X} = 3.698 with a standard deviation of S = 0.601. From Table 1, for n = 20 samples, the appropriate tolerance factor, K, is 2.396. Using the mean, standard deviation, the tolerance factor and equations 1 and 2, we can calculate an upper Tolerance Limit for Ba⁺²:

$$TL_{Upper} = \bar{X} + KS$$
 $\ln(TL_{upper}) = 3.698 + (2.396 \times 0.601)$
 $\ln(TL_{upper}) = 5.138$
 $TL_{upper} = e^{\ln(TL_{upper})} = e^{5.138} = 170 \,\mu g/L$

Within the county, the calculated upper Tolerance Limit can be used as a simple screening test for ground-water samples possibly contaminated by oilfield brines. Any barium analysis equal to or greater than an upper Tolerance Limit of $170~\mu g/L$ is considered statistically significant. If the sampled water also contains high chloride levels, for that well, oilfield brines are indicated as a probable source of contamination. As a result, water samples demonstrating elevated chloride levels and statistically significant results using the barium concentration upper Tolerance Limit are candidates for detailed hydrogeological evaluations.

GRITS/STAT

GRITS/STAT² is a freeware program from the U.S. Environmental Protection Agency for personal computers. Based upon the USEPA Statistical

Guidance Documents^{3,5}, GRITS/STAT can be used to evaluate large sets of data in a short period of time. For example, GRITS/STAT generated a probability plot of the natural log of the Ba⁺² concentrations (Figure 2), displayed the results from three tests for Normality (Figure 3) and calculated the Tolerance Limit on the transformed background data (Figure 4) in less than ten minutes. For each step, GRITS/STAT displayed summary statistics providing useful checks of the procedures applied to the set of data being evaluated.

The probability plot of the log transformed background barium concentrations (Figure 2) provides a graphical confirmation the Wood county data fits a log-normal distribution. Prior to generating the plot, GRITS/STAT displays a summary of the supporting statistics as a check of the internal calculations used to generate the probability plot. Along with the correlation coefficient, basic parameters (mean, standard deviation, etc.) and the ranking, cumulative probability and quantile for each sample are presented. Comparison values using an α of 0.01 and 0.05 are presented underneath the calculated correlation coefficient value. The comparison values can be used as a quick check of the statistical significance for the goodness-of-fit of a straight-line to the data.

In a single screen GRITS/STAT presents the results of three tests for normality - the Coefficient of Variation (CV), the Skewness Coefficient and the Shapiro-Wilk test (Figure 3). Each of the normality tests are run on the original as well as the log transformed data. Note the original data fails the Skewness Coefficient and the Shapiro-Wilk tests for normality while the log transformed data passes.

Using 95% coverage with a confidence factor of 95%, GRITS/STAT calculated a upper Tolerance Limit of 5.1384 based on the 20 log transformed background barium concentrations (Figure 4). GRITS/STAT's upper Tolerance Limit can be converted from its logarithmic form to its original

form using equation 2. GRITS/STAT leaves the conversion of the natural log up to the user.

CONCLUSIONS

Using publicly available data, information about the local hydrochemistry of oilfield brines, a few simple calculations and a single chemical analysis from each suspect well-site, an investigator can screen a large number of water samples for potential oilfield brine contamination. Water samples demonstrating elevated chloride levels and barium (Ba⁺²) concentrations equal to or greater than the upper Tolerance Limit for background barium concentrations indicate probable contamination of the tested aquifer(s) by oilfield brines.

While the calculations are not difficult, for large sets of data the screening process could require a large block of time. GRITS/STAT, free software from the U.S. Environmental Protection Agency, can be used to minimize the time required to evaluate large sets of data.

As mentioned initially, the described method may not provide absolute evidence of oilfield brine contamination in a fresh-water aquifer. However, it does identify conditions where oilfield brine contamination is a strong probability over other possible sources of chlorides. As a result, any statistically significant results should be confirmed with detailed hydrogeological evaluations.

NOMENCLATURE

K = Tolerance Factor

ln(x) = Natural Logarithm of x

n = Number of Samples

P = Tolerance Interval Coverage

S = Standard Deviation of the Sample

TL = Tolerance Limit

 \overline{X} = Sample Mean

Y = Tolerance Interval Confidence Factor

7

REFERENCES

- 1. Hem, John D.: Study and Interpretation of the Chemical Characteristics of Natural Water.

 3rd Edition, United States Geological Survey Water-Supply Paper 2254, United States Government Printing Office, Washington, D.C., (1985) 264.
- 2. United States Environmental Protection Agency: GRITS/STAT v4.2, A Ground Water Information Tracking System with Statistical Analysis Capability, United States Environmental Protection Agency, Office of Research and Development, Office of Solid Waste, Washington D.C., (November, 1992) EPA/625/11-91/002
- 3. United States Environmental Protection Agency: Statistical Analysis of Ground Water Monitoring Data at RCRA Facilities, Draft Addendum to Interim Final Guidance, United States Environmental Protection Agency, Office of Solid Waste, Washington D.C. Currently distributed as part of Statistical Training Course for Ground-Water Monitoring Data Analysis, (1992) EPA/530-R-93-003, available through the RCRA Docket.
- 4. Gibbons, Robert D.: "Statistical Tolerance Limits for Ground-Water Monitoring," Ground Water, Vol 29, No.4, (1991) p.563-570
- 5. United States Environmental Protection Agency: Statistical Analysis of Ground Water Monitoring Data at RCRA Facilities, Interim Final Guidance, United States Environmental Protection Agency, Office of Solid Waste, Washington D.C., (April, 1989) EPA/530-SW-89-026
- 6. Durfor, C.N., and Becker, Edith: <u>Public</u>
 <u>Water Supplies of the 100 Largest Cities in</u>
 the United States, United States Geological

- Survey Water-Supply Paper 1812, United States Government Printing Office, Washington, D.C., (1964) 364.
- 7. Durum, W.H., and Haffty, Joseph: "Implications of the Minor Element Content of Some Major Streams of the World," Geochimica et Cosmochimica Acta, Vol. 27,(1963) p.1-11
- 8. Collins, A. Gene: Geochemistry of Oilfield Waters, Elsevier Scientific Publishing Company, Amsterdam (1975) 496.
- 9. Rittenhouse, Gordon, Fulton, Robert B. III, Grabowski, Robert, and Bernard, Joseph L.: "Minor Elements in Oil-Field Waters," Chemical Geology, Vol. 4, (1969) p.189-209.
- 10. Davis, John C.: <u>Statistics and Data Analysis</u> in Geology, <u>Second Edition</u>, John Wiley and Sons, New York City (1986) 646.

Table 1
Tolerance Factors (K) For One-Sided Normal Tolerance
Intervals With Probability Level (Confidence Factor)
Y = 0.95 And Coverage P = 95%

<u>n</u>	<u>K</u>	<u>n</u>	<u>K</u>	<u>n</u>	<u>K</u>	<u>n</u>	<u>K</u>	<u>n</u>	<u>K</u>
3	7.655	17	2.486	55	2.036	325	1.792	675	1.746
4	5.145	18	2.543	60	2.017	350	1.787	700	1.744
5	4.202	19	2.423	65	2.000	375	1.782	725	1.742
6	3.707	20	2.396	70	1.986	400	1.777	750	1.740
7	3.399	21	2.371	75	1.972	425	1.773	775	1.739
8	3.188	22	2.350	100	1.924	450	1.769	800	1.737
9	3.031	23	2.329	125	1.891	475	1.766	825	1.736
10	2.911	24	2.309	150	1.868	<i>5</i> 00	1.763	850	1.734
11	2.815	25	2.292	175	1.850	525	1.760	875	1.733
12	2.736	30	2.220	200	1.836	550	1.757	900	1.732
13	2.670	35	2.166	225	1.824	575	1.754	925	1.731
14	2.614	40	2.126	250	1.814	600	1.752	950	1.729
15	2.566	45	2.092	275	1.806	625	1.750	975	1.728
16	2.523	50	2.065	300	1.799	650	1.748	1000	1.727

USEPA: Statistical Analysis of Ground Water Monitoring Data at RCRA Facilities, Interim Final Guidance, U.S. Environmental Protection Agency, Office of Solid Waste, Washington D.C., (April 1989) EPA/530-SW-89-026, p. B-9

Table 2
Water Well Analyses For Barium (Ba⁺²)
Wood County, Texas

		Date	Well		Date	Cl ⁻	Ba+2	
ΙD	State #	<u>Drilled</u>	<u>Depth</u>	<u>Use</u>	<u>Sampled</u>	mg/L	ug/L	$\ln(Ba^{+2})$
Α	3404607	9/86	500	Public Supply	11/15/91	170	28	3.3322
В	3404802	10/19/91	259	Public Supply	11/15/91	11	60	4.0943
C	3405508	2/10/75	545	Public Supply	11/20/91	9	20	2.9957
D	3406807	4/13/87	742	Public Supply	11/12/91	14	31	3,4340
E	3407705	1978	890	Public Supply	11/13/91	14	20	2.9957
F	3411604	05/85	402	Public Supply	11/18/91	22	46	3.8286
G	3412407	10/8/85	380	Public Supply	11/18/91	13	25	3.2189
\mathbf{H}	3413203	9/21/84	638	Public Supply	11/19/91	8	30	3.4012
I	3413809	3/2/88	611	Public Supply	11/21/91	285	50	3.9120
J	3414407	1/89	1114	Public Supply	11/19/91	12	31	3.4340
K	3420102	4/19/77	634	Public Supply	11/20/91	87	20	2.9957
L	3420207	6/24/88	512	Public Supply	11/20/91	35	20	2.9957
M	3420305	12/15/71	334	Public Supply	11/20/91	3	77	4.3438
N	3420607	6/12/87	670	Public Supply	11/20/91	109	55	4.0073
Ο	3421502	10/30/80	619	Public Supply	11/21/91	8	78	4.3567
P	3422205	5/23/83	890	Public Supply	11/21/91	11	20	2.9957
Q	3431206	10/44	510	Institution	08/13/93	7	44.7	3.8000
R	3414206	Unknown	28	Dewater	08/12/93	14	137	4.9200
S	3504402	Unknown	50	Dewater, Stock	08/11/93	6	91.2	4.5131
T	3414408	Unknown	60	Irrigation	08/12/93	26	80.4	4.3870

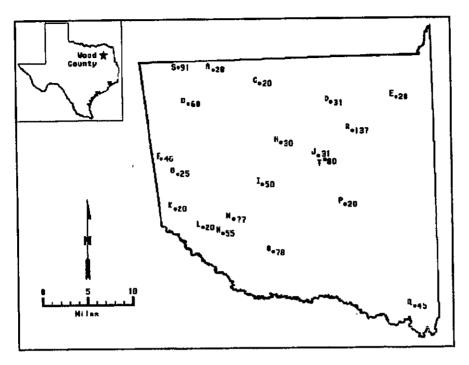


Figure 1. Background barium (Ba+2) concentrations, Wood County, Texas.



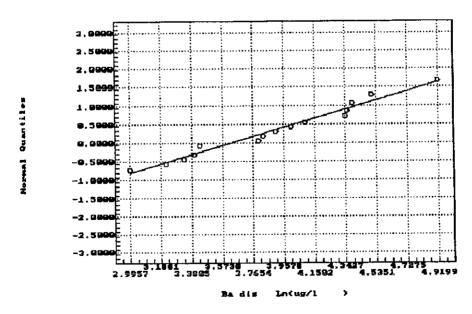


Figure 2. From GRITS/STAT, probability plot of natural log of background barium (Ba⁺²) concentrations, Wood County, Texas.

	——Summary of ———Scale		ta Norwality Observations: 28
Statistic	Original	Log	
			Hells:Background
Hean	48.2150	3.6961	Test:Observations
Std. Dev.	31.2376	8.6811	
Murtosis	1.1912	-1.0490	Shapiro-Hilk Critical Values
Minimum	20.9990	2.9957	
Haximum	137.8880	4.9200	1×:0.8688
cv	0.6479	0.1626	5x : 0 . 9858
	Scale	·	<u> </u>
Normality Test	Original	Log	Conments
Skewness Coefficient	1.2675₩	8.3475	∺ = Data may not be normally distributed
Shapiro-Hilk Test Statistic	Ø.8403×	8.9113	x ≡ Data are not normally dist- ributed at 5x significance leve
	CON	TINUE PRIN	T
			315849 Free

Figure 3. From GRITS/STAT, Normality test results for background barium (Ba⁺²) concentrations, Wood County, Texas.

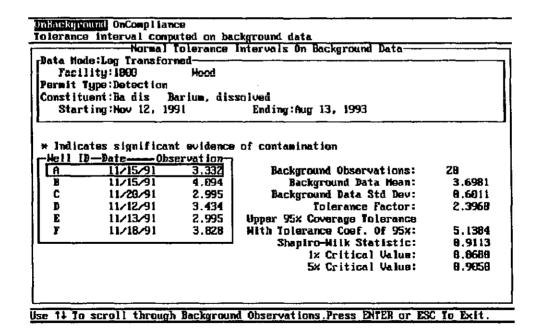


Figure 4. From GRITS/STAT, calculated Tolerance Limits on background barium (Ba⁺²) concentrations, Wood County, Texas.

